

Designing for Friction: An Autoethnographic Study of AI Sensemaking

Behnoosh Mohammadzadeh,
Université Paris-Saclay | Sorbonne Université

Main theme(s): Reflective Human-AI Interaction, Sensemaking Tools, Friction

Target domain(s): Human-Centered AI, AI Fairness

Cognitive 'target(s)': Critical reflection, Sensemaking, Narrative construction

Type of contribution & main idea

Conceptual and design-oriented contribution that operationalizes friction as a Tool for Thought in human-AI interaction through a reflexive image-captioning interface and first-person, autoethnographic inquiry.

I reframe friction as a site for sensemaking. The tool combines masking inputs and an affinity diagram as a reflexion workspace to help users linger with mismatch, trace shifts in model behavior, and construct situated narratives from lived experience. First-person methods surface subjective and affective frictions that often escape aggregate evaluations.

The AI tool's key characteristics

- Model: A BLIP image captioning model in Hugging Face, which generates a textual description each time an image is submitted or modified.
- Curated personal dataset: Users upload and select images from their own archive, grounding the interaction in personally material.
- Mask-based probing: Users can manually mask specific regions of an image using a brush tool. Each masking action triggers the model to generate a new caption, allowing users to observe how small visual absences shift the description and compare them side by side.
- Affinity canvas workspace: Image-caption pairs can be dragged onto a zoomable canvas as "instax" cards. Users rearrange, cluster, and annotate these cards with sticky notes to record observations, hypotheses.

Interaction & outcomes

How people use and interact with the AI tool for a specific activity (first column) and what outcomes this provides, enables, or leads to (second and third column).

Objective value

The 'objective' quality of the process and the outcomes – i.e. what makes them 'good' or desirable (e.g. critical thinking, understanding, new insights, learning).

Perceived value & UX

How the process of using the AI tool and its outcomes is experienced by users (e.g. how they experience the cognitive work they are doing).

Interaction design/usage strategy

Users upload a personal image and get an initial caption from a BLIP model. They mask regions with a brush tool; each mask triggers a new caption, producing multiple variants for the same image. Variants are dragged onto a zoomable affinity canvas as "instax" cards, where users place them side by side, cluster them, and add sticky-note annotations to build interpretations over time.

The tool requires deliberate user intervention (masking specific regions) before generating new outputs. Captions only change in response to explicit perturbations, preventing passive acceptance of results and structuring active probing of model behavior.

Masking feels like a way of "talking back" to the model through a shared language. Instead of accepting the first caption, I can intervene and see what changes when I remove a region. The pace is slower than typical AI use, but it pushes me to linger and probe rather than move on.

Intermediary task outcomes

After several iterations, the user has a set of masked image variants and their corresponding BLIP captions. These image-caption pairs accumulate on the affinity canvas as instax cards to show how removing a specific region shifts the description. Sticky notes capture immediate observations and questions, and rough clusters begin to form around recurring patterns (e.g., who is recognized, what is ignored, which cues trigger gendered labels).

By generating multiple caption variants tied to specific masking actions and preserving them on the canvas, the system creates a traceable record of deliberate shifts. This supports systematic comparison across variations and makes patterns of recognition and omission inspectable.

As variants accumulate, I start to feel patterns emerging. Some caption shifts are surprising; others are irritating. The canvas makes the process feel tangible because I can see the trace of what I tried, what I noticed, and how my interpretations changed.

Final task outcomes

After repeated masking and comparison, the user constructs a documented narrative explaining how the model distributes attention and recognition. The affinity canvas contains clustered image-caption variants and annotations that make visible patterns of misrecognition, omission, or bias, and their interpretation. The final outcome is a situated account of the model's behavior grounded in user's situated knowledge.

The accumulated artifacts and annotations produce a documented, revisitable account of how the model distributes attention and recognition. Instead of isolated incidents, user develops a structured explanation grounded in observable variation.

At the end, I had a situated account of what the model tends to notice, ignore, or misrecognize. At the same time, working with personal images can be emotionally heavy, because the frictions are tied to relationships and identity rather than abstract examples.

How do people's goals 'interact with' the AI tool's goals? Is there a tension?

goal:
Develop a deeper understanding of how the model distributes attention, recognition, and meaning. The tool slows down interaction and requires active probing (masking, comparison, annotation) instead of providing a definitive answer or correction.

There is a potential tension for users who want fast validation, correction, or a "better" caption. The tool does not optimize outputs or fix the model; instead, it makes friction visible and asks the user to stay with it. This can feel effortful, especially when working with personal or identity-related images, but it supports reflective sensemaking rather than efficiency.

How do you expect people to continue using this AI tool?

Continued use depends on whether users find value in the reflective process. The tool is unlikely to be used daily as a productivity aid. Instead, it may be revisited when a particular image or mismatch feels puzzling, uncomfortable, or worth investigating. To sustain engagement, the tool could support ongoing archives, allowing users to build longitudinal collections of cases, revisit past clusters, and compare patterns across time rather than treating each interaction as isolated.

What you would like to discuss

- I would like to discuss how friction can be more precisely operationalized as a Tool for Thought across different AI systems and generalized to other generative systems, including LLMs and multimodal models. I am interested in identifying transferable interaction principles that make model behavior inspectable and narratively interpretable.
- and to discuss how first-person methods can be more fully integrated into TTF frameworks: how subjective experience can be treated as epistemically valuable in designing human-AI interaction.

What would you like to take away from the workshop?

Key references (e.g. of main theories, empirical evidence, measurement methods etc.)

- [1] Angel Alexander Cabrera, Mario Tuiño Ribera, Brigitte Lee, Robert Devine, Adam Peter, and Steven M Drucker. 2023. What did my AI learn? How do data scientists make sense of model behavior. *ACM Transactions on Computer-Human Interaction* 30, 1 (2023), 1–27.
- [2] Audrey Desjardins, Oscar Tomaco, Andrés Lucero, María E Cecchinato, and Carman Neustaedter. 2021. Introduction to the special issue on first-person method(H) HCI.
- [3] Patricia Garcia and Marika Chou. 2019. Expanding our reflexive toolbox: Collaborative possibilities for examining socio-technical systems using autoethnography.
- [4] Donna Haraway. 2015. Situated knowledges: The science question in feminism and the privilege of partial perspective. 1. In *Women, science, and technology Routledge*, 455–472.
- [5] Peter Profitt and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis.
- [6] Gonzalo Ramos, Christopher Meek, Patricia Simard, Jina Suh, and Sourabh Ghoshal. 2020. Interactive machine teaching: a human-centered approach to building machine-learning models. *Human-Computer Interaction* 35, 5-6 (2020), 413–451.
- [7] Minna Ruckenstein. 2023. The feel of algorithms. *Univ of California Press*.
- [8] Lew Tansikewitch, Elena L Gassman, Jessica Ye, Aniket Kulkarni, Mira Lee, Sidra Palani, Advait Sarkar, Gonzalo Ramos, Yvonne Rogers, and Tari Subramanyam. 2025. Understanding, Processing, and Augmenting Human Cognition with Generative AI: A Synthesis of the CHI 2025 Tools for Thought Workshop. *arXiv preprint arXiv:2508.21036* (2025).

How to proceed with this work/idea?

Extend the tool beyond single-user autoethnographic use by exploring collaborative, collective and longitudinal versions of the affinity canvas, where multiple situated perspectives can be layered and built upon. Investigate how friction based interaction (masking, controlled perturbations, variant tracing) can be adapted to other generative systems such as LLMs or multimodal models.